

学校编码: 10384
学号: 23020061152472

分类号_____密级_____
UDC_____

厦 门 大 学

硕 士 学 位 论 文

大词汇量连续语音识别的性能优化研究

The Performance Optimization Research on Large Vocabulary Continuous Speech Recognition

欧建林

指导教师姓名: 史晓东 教授

专 业 名 称: 计算机应用技术

论文提交日期: 2009 年 月

论文答辩时间: 2009 年 月

学位授予日期: 2009 年 月

答辩委员会主席: _____

评 阅 人: _____

2009 年 月

厦门大学博硕士论文摘要库

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学博硕士论文摘要库

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

厦门大学博硕士论文摘要库

摘要

大词汇量连续语音识别 (Large Vocabulary Continuous Speech Recognition, LVCSR) 是语音识别研究的重点和难点之一, 其涉及了声学模型、语言模型、搜索算法等多方面的知识和技术; 本文在介绍语音识别的基本原理基础上, 就如何提高 LVCSR 的识别速度以及识别准确率两方面进行了研究和探讨。

在基于 HMM 的 LVCSR 系统中其声学建模常用连续密度 HMM 来实现, 每个 HMM 状态都表示为一个高斯混合模型 (GMM), 而每个 GMM 模型又包含较多的高斯分量, 这使得状态似然率的计算量非常大, 是语音识别速度慢的主要原因之一, 因此有必要设计有效的似然率快速算法在不降低或不明显降低识别准确率的前提下加快似然率的计算。本文首先分析了 HMM 中的似然率计算, 并分析了采用并行方式实现似然率计算的可行性, 在此基础上提出了一种基于 SIMD 的似然率并行计算算法, 并利用 HTK3.4 工具包作为实验基线系统、TIMIT 和 WSJ0 语料库作为实验语料库搭建实验平台, 将此算法与部分距离消去算法 (PDE)、最佳混合分量预测算法 (BMP)、特征矢量重排算法 (FCR) 以及高斯选择算法 (GS) 等其他经典快速似然率算法进行比较; 实验结果表明, 该算法在不降低识别准确率的前提下显著降低似然率计算开销, 并且性能优于其他几种似然率快速算法。

为了将语义信息与 N-gram 统计语言模型结合并运用到语音识别过程中以提高 LVCSR 系统的识别准确率, 本文研究了潜在语义分析 (LSA) 理论及其在 LVCSR 系统中应用的相关技术, 在此基础上利用 WSJ0 文本语料库构建 LSA 模型, 并将其与 N-gram 模型进行插值组合, 构建了包含语义信息的混合模型; 同时为了优化混合模型的性能, 利用基于密度函数初始化类中心的 K 均值聚类技术对 LSA 模型的向量空间进行聚类, 并提出平滑计算方法对概率进行平滑。WSJ0 语料库上的模型困惑度实验和连续语音识别实验结果表明: 混合模型性能优于 N-gram, LSA 能在一定程度上辅助 N-gram 提高 LVCSR 的识别率。

关键词: 似然率快速计算; 潜在语义分析; 大词汇量连续语音识别

厦门大学博硕士论文摘要库

Abstract

Large vocabulary continuous speech recognition (LVCSR) is one of the most important subjects of spoken language processing, which involves many knowledge sources and techniques such as acoustic model, language model and decoding algorithm. This paper will introduce the basic knowledge of speech recognition and then discuss how to improve the real-time performance of speech recognition systems and how to improve the recognition accuracy.

Most LVCSR systems are based on statistical models, which use continuous density HMM as the underlying technology to perform acoustic modeling of speech signals. In this system, each state is a Gaussian mixture model (GMM) which is consisted of many Gaussian mixtures. For this kind of likelihood-based speech recognition systems, the state likelihoods estimation is computationally intensive. It is one of the most important reasons why the recognition is so slow. Therefore it is necessary to develop efficient techniques in order to reduce the computational overhead of likelihood computation without any degradation or a significant degradation of recognition accuracy. The likelihood computation of LVCSR systems which are based on continuous density HMM is analyzed to show that the conventional way of sequential computing is time-consuming and the likelihood computation itself can be implemented in parallel. A SIMD-based algorithm which can carry out parallel likelihood computation is presented in this paper. By taking HTK 3.4 toolkit as the baseline system and TIMIT、WSJ0 corpus as the experiment corpus, the experiment platform is built. And then the algorithm is compared to other efficient techniques such as partial distance elimination (PDE), best mixture prediction (BMP), and feature component reordering (FCR) and Gaussian selection (GS) on this platform. Experiments results show that the SIMD-based algorithm can significantly reduce the time overhead of likelihood computation without any degradation of recognition accuracy. And the performance is better than other fast computation techniques.

In order to integrate the semantic knowledge with N-gram language model for LVCSR to improve recognition accuracy, the theory of latent semantic analysis (LSA) and the related techniques for applying it in LVCSR system is described in this paper. And then LSA model is constructed on the WSJ0 text corpus. We use the interpolation method to combine this model with conventional N-gram to form a hybrid language model which include semantic knowledge. To optimize the performance of the hybrid model, we apply k-means algorithm to perform vector clustering in the LSA vector space while the density function is used to initialize the centroids, and propose a computation method for smoothing the probabilities. The model perplexity tests and continuous speech recognition experiments are conducted on the WSJ0 corpus. Results show that the constructed hybrid language model outperforms the corresponding N-gram and can improve the recognition of LVCSR to some extent.

Key Words: Fast Likelihood Computation; Latent Semantic Analysis; Large Vocabulary Continuous Speech Recognition

目 录

第一章 绪论	1
1.1 语音识别的发展	1
1.2 语音识别的分类	2
1.3 本文研究背景和目的	3
1.4 本文的主要工作与结构	4
第二章 语音识别基本理论	5
2.1 预处理	5
2.2 特征提取	7
2.3 声学模型	8
2.3.1 声学单元选取.....	8
2.3.2 HMM基本概念.....	9
2.3.3 HMM训练算法.....	11
2.4 语言模型	14
2.4.1 N元文法模型.....	14
2.4.2 结构化语言模型.....	16
2.5 识别算法	17
2.5.1 Viterbi-Beam算法.....	17
2.5.2 Token-Passing算法	18
2.6 识别结果评测	19
2.7 本章小结	20
第三章 基于SIMD的似然率快速计算.....	21
3.1 HMM的似然率计算.....	21
3.2 似然率快速算法	23
3.2.1 部分距离消去算法.....	23
3.2.2 最佳分量预测算法.....	25
3.2.3 特征矢量重排算法.....	25
3.2.4 高斯选择算法.....	26

3.3 基于SIMD的并行算法	27
3.3.1 Intel SIMD技术	27
3.3.2 算法设计	30
3.4 实验与结果分析	33
3.5 本章小结	37
第四章 潜在语义分析在LVCSR中的应用	39
4.1 潜在语义分析基本原理	39
4.1.1 特征表示	40
4.1.2 奇异值分解	42
4.1.3 概率计算	43
4.2 潜在语义分析与N-gram的组合	44
4.3 聚类平滑技术	45
4.3.1 聚类算法设计	45
4.3.2 平滑技术	47
4.4 语义模型的应用方法	49
4.5 实验与结果分析	51
4.5.1 N-gram模型的建立	51
4.5.2 潜在语义分析模型的建立	52
4.5.3 连续语音识别实验	53
4.6 本章小结	54
第五章 总结与展望	57
参考文献	59
研究生期间发表的论文	63
致 谢	65

Contents

Chapter 1 Introduction.....	1
1.1 Development of Speech Recognition	1
1.2 Classification of Speech Recognition.....	2
1.3 Research Background and Objective	3
1.4 Contents and Structure.....	4
Chapter 2 Basic Principle of Speech Recognition.....	5
2.1 Pre-processing	5
2.2 Feature Extraction	7
2.3 Acoustic Model	8
2.3.1 Selection of Acoustic Units.....	8
2.3.2 Definition of HMM.....	9
2.3.3 Training Algorithms of HMM.....	11
2.4 Language Model.....	14
2.4.1 N-gram Language Model.....	14
2.4.2 Structured Language Model.....	16
2.5 Decoding Algorithms	17
2.5.1 Viterbi-Beam Algorithm	17
2.5.2 Token-Passing Algorithm.....	18
2.6 Evaluation of Results	19
2.7 Conclusion	20
Chapter 3 Using SIMD to Speed up Likelihood Computation.....	21
3.1 Likelihood Computation of HMM	21
3.2 Algorithms of Fast Likelihood Computation.....	23
3.2.1 Partial Distance Elimination Algorithm.....	23
3.2.2 Best Mixture Prediction Algorithm.....	25
3.2.3 Feature Component Reordering Algorithm	25

3.2.4 Gaussian Selection Algorithm.....	26
3.3 SIMD-Based Parallel Algorithm.....	27
3.3.1 Intel SIMD Technology	27
3.3.2 Design of the Algorithm.....	30
3.4 Experiments and Result Analysis	33
3.5 Conclusion	37
Chapter 4 Application of Latent Semantic Analysis in LVCSR	39
4.1 Basic Principle of Latent Semantic Analysis	39
4.1.1 Feature Presentation.....	40
4.1.2 Singular Value Decomposition.....	42
4.1.3 Computation of Probabilities	43
4.2 Combining Latent Semantic Analysis with N-gram	44
4.3 Smoothing Technology of Clustering	45
4.3.1 Design of the Clustering Algorithm.....	45
4.3.2 Smoothing Technology	47
4.4 The Method of Applying Latent Semantic Analysis Model.....	49
4.5 Experiments and Result Analysis	51
4.5.1 Construction of N-gram Model.....	51
4.5.2 Construction of Latent Semantic Analysis Model	52
4.5.3 Experiments of Continuous Speech Recognition	53
4.6 Conclusion	54
Chapter 5 Conclusion and Future Work	57
References	59
Published Papers.....	63
Acknowledgement.....	65

第一章 绪论

语音识别技术是让机器通过识别和理解过程把语音信号转变为相应的文本或命令的高技术；其以语音信号为研究对象，涉及语言学、计算机科学、信号处理、生理学、心理学等诸多领域，是模式识别的重要分支，也是人机接口设计的一项重要内容，其最终目标是实现人与机器的自然语言通信。

如今，语音识别几乎可以应用于人们日常生活的各个领域，目前已经在金融、旅游等行业的语音咨询，工业生产部门的语音控制，声控智能玩具，电话、电信系统的自动拨号以及辅助控制与查询等方面获得成功的应用，并且有望成为下一代操作系统和应用程序的用户界面。随着信息产业和互联网的迅速发展，对语音识别的需要也更加紧迫，因此研究语音识别，开发相应的产品有着广泛的社会意义和经济意义^[1,2]。

作为智能计算机研究的主导方向和人机通信的关键技术，语音识别一直受到各国学界的广泛关注。现在不仅有以三个国际语音大会（ICASSP、ICSLP、Euro Speech）为代表的各类国际会议论坛为语音研究者提供交流的舞台，有IEEE Trans. Speech and Audio Processing、Speech Communication等为代表的权威杂志登载语音领域的最新进展，有HTK^[3]、Sphinx^[4]、Julius^[5]等为代表的优秀实验平台、有WSJ^[6]、TIMIT^[7]为代表的优秀语料库，还有很多成熟的商业实用系统，如：IBM的ViaVoice、美国Dragon公司的Naturally-Speaking等^[8]。

1.1 语音识别的发展

语音识别的研究工作开始于 50 年代，当时 AT&T Bell 实验室实现了第一个可识别十个英文数字的语音识别系统——Audry 系统，这是语音识别研究工作的真正开端。

到六十年代中期，由于计算机的发展提供了实现复杂算法的软、硬件环境，以及数字信号处理的理论和算法的进展提供了语音识别的技术基础，大规模的语音识别研究工作得以展开，并且逐步取得了实质性的进步。其中，最重要的发展是语音信号线性预测编(LPC)技术和动态时间规整(DTW)技术的提出，它们

有效地解决了语音信号的特征提取和不等长匹配问题，对特定人的语音识别十分有效。

80 年代以来，统计方法逐步取代了传统的模板匹配方法，成为语音识别技术的主流。统计方法不再追求语音特征的细化，而是更多地从整体平均(统计)的角度来建立最佳的语音识别系统。隐马尔可夫模型(HMM)就是其中的典型代表，该方法可以在音素、半音节、音节、词、短语甚至句子中的任一级建立相应的随机统计模型，并且可以把语言模型结合到识别过程中，提高识别的准确率。统计方法的应用有效地把特定人、小词汇量、孤立词或连接词的识别，发展到了非特定人、大词汇量、连续语音识别。1988 年，美国 CMU 大学用 VQ/HMM 的方法实现了 Sphinx，这是一个高性能的非特定人、大词汇量连续语音识别系统。随着语音识别技术的进一步成熟，许多商用大词量连续语音识别系统被推出，例如 IBM 的 ViaVoice，微软的 Whisper。

近年来，中文的语音识别取得很大的发展，我国语音识别研究工作一直紧跟国际水平、大词汇连续语音识别的研究列入了国家“863”计划，由中科院声学所、自动化研究所、北京大学等单位研究开发。同时鉴于中国未来的庞大市场，国外的研究机构和公司也非常重视中文的语音识别研究和开发，IBM、微软、Dragon 等国际大公司相继投入到汉语语音识别系统的开发中，其投资也逐年增加，产品逐渐走向市场^[9,10]。

1.2 语音识别的分类

从不同的角度和要求出发，语音识别有不同的分类方法^[11]。

按照词汇表大小分，有小词汇量（词汇量小于 100）、中词汇量（词汇量在 100 到 1000 之间）、大词汇量（词汇量在 1000 以上）。

按照语音发音方式来分，有孤立词识别、连接词识别、连续语音识别 3 种；所谓孤立词识别（Isolated Word Recognition）是指在发待识别音时，每次只含词汇表中的一个词条；连接词识别（Connected Word Recognition）是指每次说词汇表中的若干词条来进行识别，该若干词条以慢速连续的方式连续说出，一般指 0~9 十个数字连接而成的多位数字的识别，并包含其他一些少量的操作指令等；连续语音识别（Continuous Speech Recognition）指说话人以自然的方式

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库